MICROBIOLOGY

Nitrogen-fixing organelle in a marine alga

Tyler H. Coale[1]*†, Valentina Loconte[2,3]*†, Kendra A. Turk-Kubo[1], Bieke Vanslembrouck[2,3], Wing Kwan Esther Mak[1], Shunyan Cheung[4], Axel Ekman[2], Jian-Hua Chen[2,3], Kyoko Hagino[5], Yoshihito Takano[5], Tomohiro Nishimura[6,7], Masao Adachi[7], Mark Le Gros[2,3], Carolyn Larabell[2,3], Jonathan P. Zehr[1]*

Symbiotic interactions were key to the evolution of chloroplast and mitochondria organelles, which mediate carbon and energy metabolism in eukaryotes. Biological nitrogen fixation, the reduction of abundant atmospheric nitrogen gas (N_2) to biologically available ammonia, is a key metabolic process performed exclusively by prokaryotes. *Candidatus* Atelocyanobacterium thalassa, or UCYN-A, is a metabolically streamlined N_2-fixing cyanobacterium previously reported to be an endosymbiont of a marine unicellular alga. Here we show that UCYN-A has been tightly integrated into algal cell architecture and organellar division and that it imports proteins encoded by the algal genome. These are characteristics of organelles and show that UCYN-A has evolved beyond endosymbiosis and functions as an early evolutionary stage N_2-fixing organelle, or "nitroplast."

Symbioses, the living together of two or more species, involve varied metabolic interactions and diverse eukaryotic and prokaryotic organisms (*1*). Such interactions between unicellular organisms were key for evolution of the organelles that are the basis of eukaryotic life on Earth, including the metabolic innovations of aerobic respiration and photosynthesis of mitochondria and chloroplasts, respectively (*2*, *3*). Despite the success of such organelles, primary endosymbiosis remains exceptionally rare, with only three occurrences ever described (*4*). Biological nitrogen (N_2) fixation, the reduction of abundant inert atmospheric N_2 gas to ammonia, is a key metabolic innovation that maintains the fertility of terrestrial and aquatic environments. N_2 fixation in eukaryotes is only known to exist through diverse symbioses with Bacteria or Archaea, the microbes capable of N_2 fixation (*5*). Several presumed endosymbiotic interactions have been described that involve unicellular eukaryotic algae and N_2-fixing intracellular spheroid bodies that are evolutionarily related to cyanobacteria (*6*), including diatom symbionts (*7*) and the *Candidatus* Atelocyanobacterium thalassa (UCYN-A) symbiosis with a haptophyte alga (*8*). However, it has remained unclear where these symbiotic relationships fall along the spectrum of endosymbiosis to organelle, and no N_2-fixing organelle has yet been discovered (*9*, *10*).

The UCYN-A symbiosis is between closely related strains of a metabolically streamlined N_2-fixing cyanobacterium (UCYN-A) and the marine haptophyte algae *Braarudosphaera bigelowii* and relatives (*8*, *11*, *12*). Different strains include coastal and open-ocean ecotypes (*13*) that are widely distributed throughout the global ocean, including polar waters (*14*), and are important in global marine N_2 fixation (*15*). Both ecotypes have been shown to have some characteristics of an organelle, including metabolic streamlining as a result of genome reduction, evidenced by the loss of carbon fixation and oxygenic photosynthesis genes, and the entire tricarboxylic acid cycle (*16*). It has also recently been shown that the sizes of the partners in the different strains have an organelle-like size relationship (*17*). However, key biological features of organelles, such as coordinated cell division, protein import by the endosymbiont, or gene exchange between partner cells, as have been shown in the plastid-like chromatophore of the protist *Paulinella* (*4*), have never been demonstrated in UCYN-A or any of the N_2-fixing spheroid bodies in diatoms (*9*). A recently obtained stable culture of *B. bigelowii*/UCYN-A (the coastal strain, UCYN-A2) was used to interrogate UCYN-A for organelle-like characteristics. Here we present results of soft x-ray tomography (SXT) (*18*) and proteomic analyses showing that (i) UCYN-A is integrated into the *B. bigelowii* cell architecture and divides in a defined synchronous sequence with the organelles and (ii) it receives proteins that are encoded by the *B. bigelowii* genome and targeted to UCYN-A, both of which are criteria commonly used to distinguish organelles from endosymbionts (*19*). Thus, UCYN-A appears to have molecular and cellular characteristics of an early stage N_2-fixing organelle.

B. bigelowii cell architecture and organelle division

UCYN-A is a membrane-enclosed spherical body [which is enclosed in an internal *B. bigelowii* membrane (*12*)] that lies at the posterior end of the *B. bigelowii* cell, opposite the end with the haptonema and between the two chloroplasts (Fig. 1, A and B, and fig. S3A). UCYN-A cannot fix CO_2, as it has lost the genes for carbon fixation and oxygenic photosynthesis, but has a complete set of *nif* genes and fixes N_2 (*16*, *20*). Isotopic tracer studies showed that fixed N is rapidly transferred to the *B. bigelowii* cell, and *B. bigelowii* fixes CO_2 into metabolites that are taken up by UCYN-A (*8*). The intracellular location of UCYN-A with respect to the haptophyte organelles is likely to be critical for the exchange of numerous metabolites (*21*) and regulation of oxygen flux given that UCYN-A is known to only fix N_2 during the day when the haptophyte is evolving oxygen (*22*) and the N_2-fixing enzyme, nitrogenase, is sensitive to inactivation by oxygen (*23*). There is only a single UCYN-A per *B. bigelowii* cell, so division of UCYN-A must be coordinated with the *B. bigelowii* eukaryotic cell to maintain the unicellular symbiosis (*24*).

We determined the three-dimensional (3D) structure of *B. bigelowii* and dynamics of intracellular body division with SXT (Fig. 1 and fig. S1). Prior imaging studies used 2D techniques, such as transmission electron microscopy (*11*, *12*), that did not resolve the precise location of UCYN-A relative to the organelles or determine how UCYN-A divides relative to other *B. bigelowii* cell constituents. SXT not only resolves the 3D structure of cellular components but also measures changes in their density and composition using the linear absorption coefficient (LAC) (*25*). Using SXT, hundreds of intact individual specimens were imaged, and 49 were segmented for volume and LAC measurements to quantify cell cycle changes in the structural organization and molecular density of cellular components throughout the daily light-dark cycle (see supplementary methods).

UCYN-A is a prominent intracellular body (15.2 ± 3.3 μm^3) which occupies 8.6% of the intracellular space (Fig. 1A and tables S1 and S3). The mitochondrial network wraps almost entirely around UCYN-A and separates it from the nucleus (fig. S2A and movie S1). Quantitative analysis of the 3D SXT data showed that the basic structural organization remains relatively constant throughout the light cycle, when UCYN-A is fixing N_2 (Fig. 1A and fig. S3C) (*22*). The positioning of mitochondria around UCYN-A is similar to mitochondrial–organelle interactions seen in other cells at times of increased energy demand, such as in mammalian neurons and pancreatic cells (*26*, *27*). The proximity and arrangement of these structures suggests that the large adenosine triphosphate demand of N_2

---

[1] Ocean Sciences Department, University of California, Santa Cruz, CA, USA. [2] Department of Anatomy, School of Medicine, University of California, San Francisco, CA, USA. [3] Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [4] Institute of Marine Biology and Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan. [5] Marine Core Research Institute, Kochi University, Nankoku, Kochi, Japan. [6] Fisheries Technology Institute, Japan Fisheries Research and Education Agency, Hatsukaichi, Hiroshima, Japan. [7] Laboratory of Aquatic Environmental Science, Faculty of Agriculture and Marine Science, Kochi University, Nankoku, Kochi, Japan.

*Corresponding author. Email: tcoale@ucsc.edu (T.H.C.); valentina.loconte@ucsf.edu (V.L.); zehrj@ucsc.edu (J.P.Z.)

†These authors contributed equally to this work.

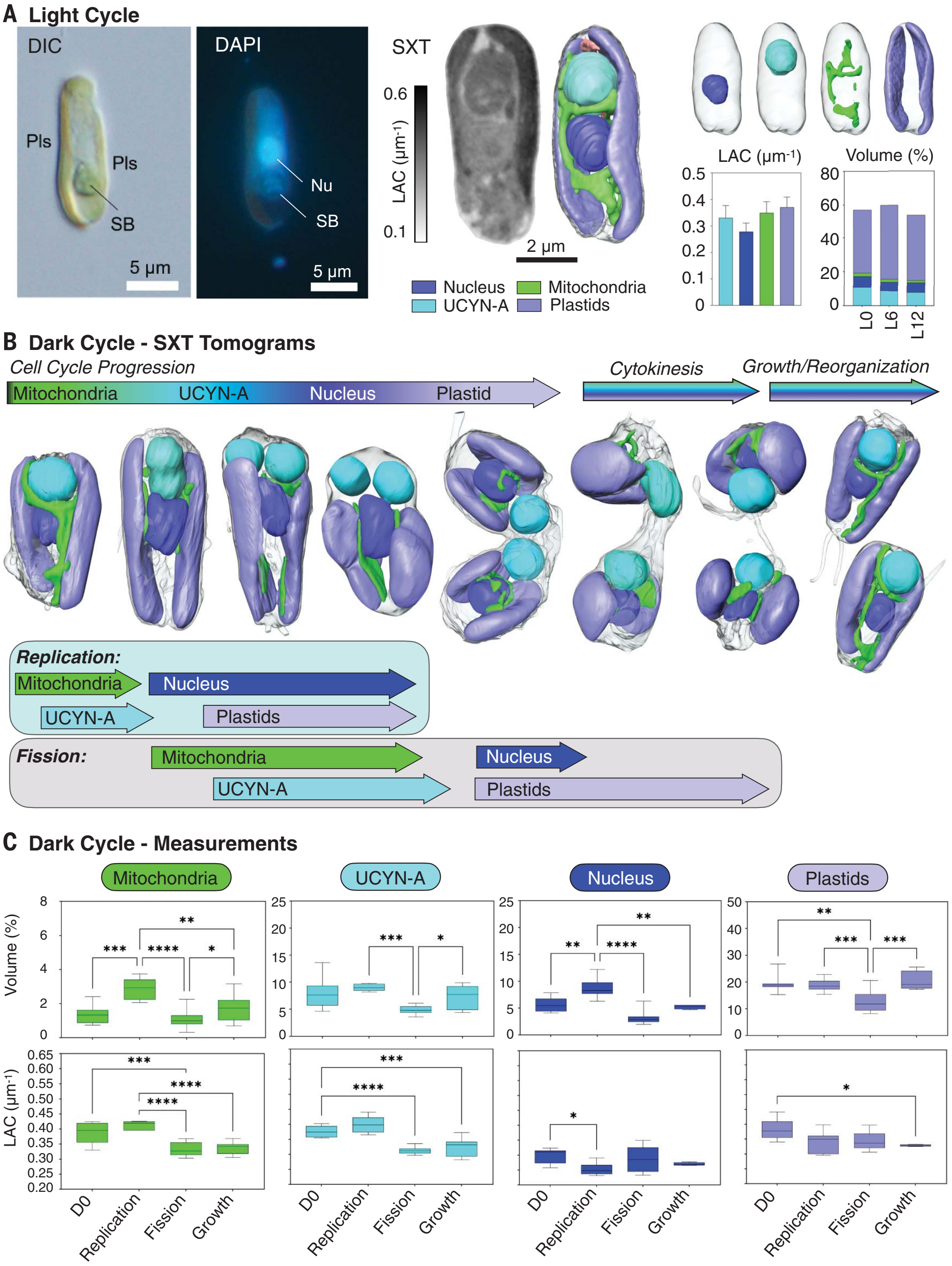

Fig. 1. *B. bigelowii*/UCYN-A light-dark cycle is highly coordinated. (**A**) Structural characterization during the light cycle: (Left) Differential interference contrast (DIC) micrograph displays the whole alga and showcases UCYN-A spheroid body (SB) and plastids (Pls) in the FR21 culture; the correspondent 4′,6-diamidino-2-phenylindole (DAPI) staining reveals the cell nucleus (Nu). In both images, the UCYN-A spheroid body (SB) is identified above the nucleus. Scale bars, 5 μm. (Middle) Orthoslice from SXT data; color-coded 3D renderings of subcellular components. (Right) LAC (molecular density) and volume measurements of UCYN-A and organelles. (**B**) 3D SXT renderings at intervals during the dark cycle (left to right) depict the replication and fission of UCYN-A and *B. bigelowii* organelles; quantification of changes in volumes and LAC measurements are shown in (**C**). The first detectable change was a twofold increase in mitochondrial volume between the onset [dark 0 (D0)] and midpoint of the dark cycle, followed by mitochondrial fission into two networks of equal volumes and decreased density (LAC). Next was replication and fission of UCYN-A into two smaller bodies with decreased LAC. SXT measurements revealed that replication of the host nucleus and plastids had occurred in those cells with two UCYN-A. Fission of the host nucleus and plastids was detected next, with a roughly 50% decrease in volumes. Subsequent cytokinesis produced two identical endosymbionts; growth of UCYN-A, host nucleus, and plastids proceeded throughout the rest of the dark cycle. Statistical difference was determined using a cutoff of 5% ($P < 0.05$). Differences in the organelle volume and LAC are reported as follows: mitochondrial volume, ****$P < 0.0001$, ***$P = 0.005$, **$P = 0.0035$, and *$P = 0.0188$ (Tukey's multiple comparison test); mitochondria LAC, ****$P < 0.0001$, ***$P = 0.0004$ (Tukey's multiple comparison test); UCYN-A volume, ***$P = 0.0009$ and *$P = 0.0133$ (Dunn's multiple comparison test); UCYN-A LAC, ****$P < 0.0001$ and ***$P = 0.0005$ (Tukey's multiple comparison test); nuclear volume, ****$P < 0.0001$, **$P = 0.0023$ (Tukey's multiple comparison test); nuclear LAC, *$P = 0.0105$ (Dunn's multiple comparison test); plastid volume, ***$P = 0.0008$ and **$P = 0.0011$ (Dunn's multiple comparison test); plastid LAC, *$P = 0.0396$ (Dunn's multiple comparison test).

fixation in UCYN-A is supplied by the closely associated mitochondria.

The process of mitotic cell replication (Fig. 1B), which occurs during the dark cycle, was initiated with replication of the mitochondria. The mitochondrial volume of cells examined midway through the dark cycle was twice that of cells examined at the beginning of the dark cycle (Fig. 1C). The mitochondrial network divides into two segments (fig. S2B and movie S2), each with a decreased LAC value suggesting the presence of fewer cristae and, therefore, reduced cellular respiration. Replication and fission of UCYN-A proceeded in the absence of surrounding mitochondria, which were at this point concentrated between the *B. bigelowii* nucleus and plastids. There was minimal volumetric change of UCYN-A during replication,

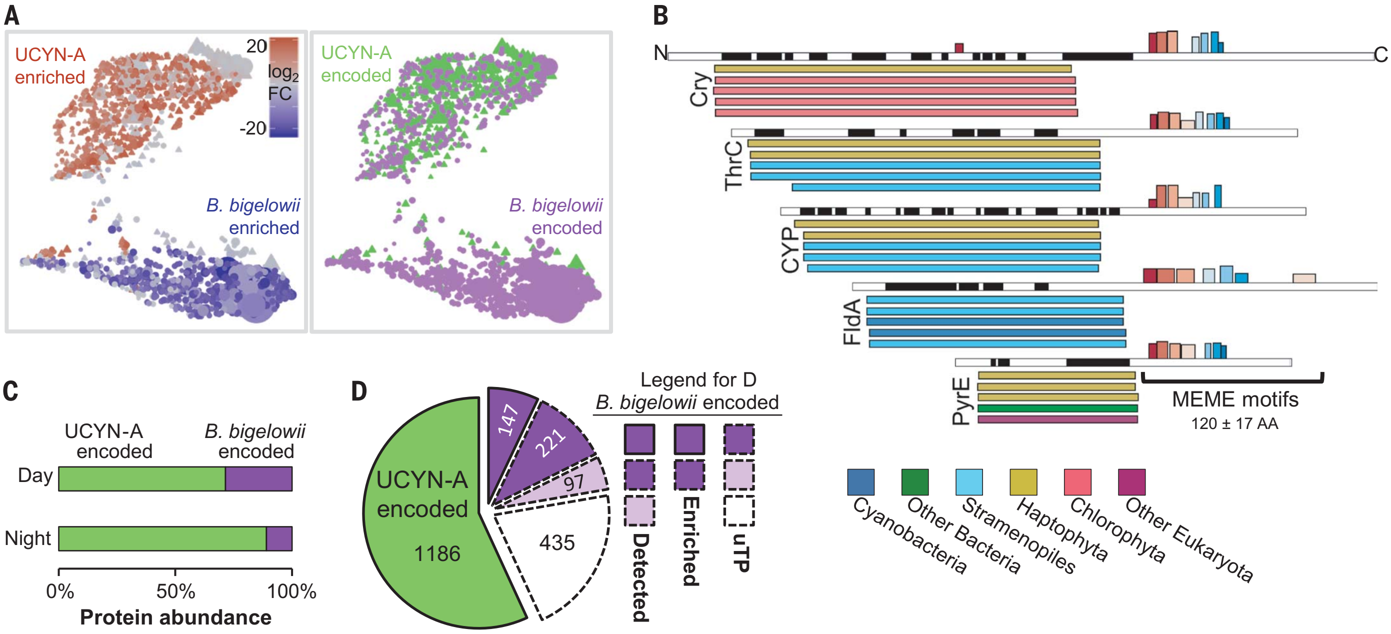

Fig. 2. *B. bigelowii* contribution to the UCYN-A proteome. (**A**) Uniform manifold approximation and projection (UMAP) clustering analysis reconstructed from all detected protein relative abundances from all samples. *B. bigelowii*–encoded proteins shown as circles, UCYN-A encoded as triangles. Icon size indicates average relative abundance in all samples. The left plot shows individual proteins colored by their average $\log_2$ fold change (FC) between UCYN-A and *B. bigelowii* samples. Right plot shows proteins colored by their genome of origin. (**B**) Selected *B. bigelowii* proteins detected in UCYN-A indicated with white bars, and black segments correspond to mass spectrometry–identified peptides. Top five BlastP hits are shown as bars colored by taxonomic origin. Motifs identified by MEME block diagram shown as colored blocks (see supplementary methods). ThrC, threonine synthase. (**C**) Average contribution to UCYN-A proteome by UCYN-A and *B. bigelowii*–encoded proteins in day and night samples. Only proteins significantly more abundant in UCYN-A samples (DESeq2 Wald test $P < 0.05$) are included. (**D**) The comprehensive UCYN-A proteome. Numbers indicate count of unique proteins in each category. "Detected" refers to identification in the proteomic dataset. "Enriched" indicates statistically more abundant in UCYN-A samples (DESeq2 Wald test $P < 0.05$). uTP indicates UCYN-A transit peptide detected by HMM search.

but fission was made obvious by the presence of numerous cells with two distinct but smaller UCYN-A bodies (9.9 ± 1.8 μm^3 and 5.0 ± 0.7%; table S2). There was also an ~18% decrease in LAC value for each UCYN-A (Fig. 1C and table S2), suggesting that the newly formed UCYN-A are less densely packed immediately after fission. As UCYN-A completed division, the *B. bigelowii* nuclei replicated, which is shown by a 37.5% increase in the percent volume (Fig. 1C and fig. S2C) and decreased LAC values, indicating that less volume is required for the replicated DNA than for the new nuclear envelope and nucleoplasm. As duplication of the nucleus continued, distinct invaginations in both *B. bigelowii* plastids were detected, indicating the onset of plastid replication and fission (Fig. 1B and fig. S2D). Fission of the plastids and *B. bigelowii* nucleus occurred simultaneously, followed by cell division (movie S3). Cell division appears to be first led by migration of the organelles and then completed with the migration of UCYN-A. The precise sequence of division of UCYN-A along with *B. bigelowii* organelles shows that UCYN-A is integrated into cell growth and division, and the coordination implies control by *B. bigelowii*, which was previously suggested by whole-genome expression analysis (*24*). Such integration is a required evolutionary step along the trajectory of evolution from an endosymbiont to an organelle (*28*, *29*). The SXT results show not only that UCYN-A is distinctly positioned with respect to organelles in the *B. bigelowii* cell but also that it participates in synchrony with *B. bigelowii* organelle and cell division, which ensures that UCYN-A is vertically transmitted to daughter cells along with the organelles.

UCYN-A imports *B. bigelowii* nuclear-encoded proteins

To investigate the contribution of *B. bigelowii* genome-encoded proteins to the UCYN-A proteome, the protein composition of UCYN-A was determined by comparing protein abundances in the whole symbiosis and isolated UCYN-A bodies. *B. bigelowii* cells and isolated UCYN-A were obtained by filtration, cell disruption, and density gradient centrifugation (fig. S3B). Purity of isolated UCYN-A was determined by light microscopy, flow cytometry, and digital droplet polymerase chain reaction targeting specific nuclear and UCYN-A markers (figs. S3B and S4, A to D), which was then corroborated by the proteomic profiling. For example, during the day, the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) protein RbcL was >2200 times more abundant in whole *B. bigelowii* samples, whereas the nitrogenase iron protein (NifH) was 37 times more abundant in UCYN-A samples (table S4). Using predicted proteins from the UCYN-A genome and *B. bigelowii* genomic and transcriptomic datasets (*12*), 1804 total proteins were detected, including 609 encoded in the UCYN-A genome and 1156 from *B. bigelowii*. Proteins that were more abundant in isolated UCYN-A relative to *B. bigelowii* (DESeq2 Wald test, $P < 0.05$) were considered enriched and assigned to the comprehensive UCYN-A proteome used in further analyses.

Most proteins encoded by either genome were more abundant in their respective sample type (data S1 and S2 and fig. S3E), but many *B. bigelowii*–encoded proteins were found in UCYN-A (Fig. 2A and fig. S4E). When all proteins significantly enriched in UCYN-A samples were weighted by their abundance, *B. bigelowii*–encoded proteins were found to make up about 28% of the UCYN-A proteome during the day and 11% at night (Fig. 2C). In total, 368 *B. bigelowii*–encoded proteins were determined to be more abundant in UCYN-A samples. The many nuclear-encoded proteins contained in UCYN-A indicate import from the surrounding *B. bigelowii* cell.

Putatively imported protein sequences had characteristics of import machinery associated

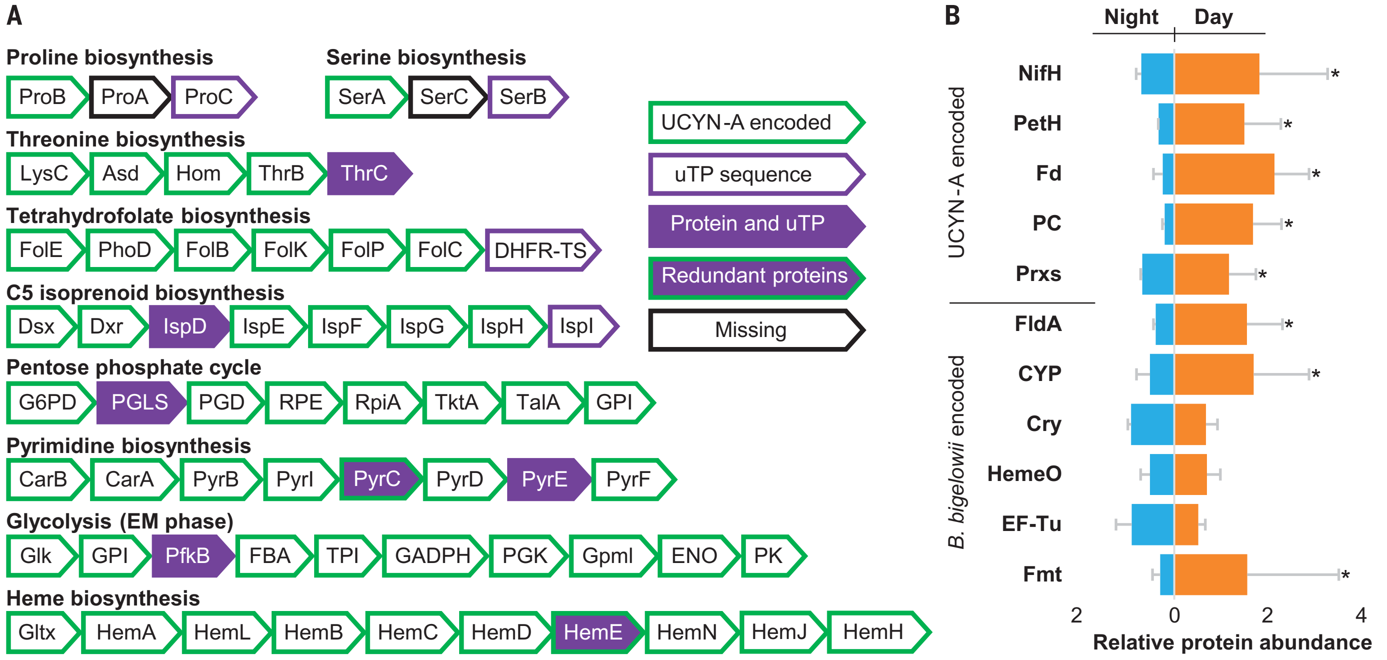

Fig. 3. *B. bigelowii* participation in UCYN-A metabolic pathways and daytime N_2 fixation. (**A**) UCYN-A and *B. bigelowii* genes in UCYN-A metabolic pathways. Green outline indicates a UCYN-A–encoded gene. Purple outline indicates a uTP sequence but no protein detected. Purple fill indicates that a *B. bigelowii* protein was identified by a significant increase in protein abundance (DESeq2 Wald test, $P < 0.05$) in UCYN-A samples and the presence of a uTP sequence. Black outline shows genes missing from the UCYN-A genome and lacking a uTP-containing *B. bigelowii* homolog. See data S7 for abbreviations. (**B**) Relative abundance of UCYN-A and uTP-containing *B. bigelowii* proteins in day and night UCYN-A samples. Error bars show ±1 SD. Asterisk indicates significant increase in daytime abundance versus night (DESeq2 Wald test, $P < 0.05$). NifH, dinitrogenase reductase; PetH, ferredoxin NADP reductase; Fd, ferredoxin; PC, plastocyanin; Prxs, peroxiredoxin; FldA, flavodoxin; CYP, cytochrome P450; Cry, cryptochrome; HemeO, heme oxygenase; EF-Tu, elongation factor thermal unstable; Fmt, methionyl-tRNA formyltransferase. Fd, FldA, CYP, and Cry show the sum of all detected proteins sharing these annotations.

with translocation to organelles. Many of the *B. bigelowii*–encoded proteins present in UCYN-A were considerably longer than orthologous proteins owing to a C-terminal extension (Fig. 2B). Alignment of these proteins revealed a series of conserved motifs within a region lacking any known protein domains (fig. S5). Analysis of proteomic mass spectrometry data failed to indicate peptides matching this C-terminal region, potentially indicating that it is cleaved or untranslated (Fig. 2B). The presence of this sequence within all predicted *B. bigelowii* proteins was determined using a custom hidden Markov model (HMM) constructed from 233 *B. bigelowii* protein sequences enriched in UCYN-A samples (fig. S5B). We detected the sequence in 258 of the 368 *B. bigelowii*–encoded proteins enriched in UCYN-A, which accounted for 62.6% of all *B. bigelowii* protein quantified in UCYN-A (fig. S6C). We also recovered 476 additional *B. bigelowii* proteins that contained this sequence but were not detected in any proteomes. It appears that this sequence is a distinctive transit peptide, analogous to those used in protein import in chloroplasts (*30*), mitochondria (*31*), and the chromatophore (*32*). This UCYN-A transit peptide (uTP) presumably directs proteins to UCYN-A and therefore may constitute a marker that can identify UCYN-A–localized proteins (table S5). Combining the detected *B. bigelowii*–encoded UCYN-A proteins with those predicted on the basis of the HMM search yielded 900 proteins potentially complementing the 1186 encoded by the UCYN-A genome (Fig. 2D). As previously reported, there is no evidence of gene transfer from UCYN-A to the *B. bigelowii* nuclear genome (*12*), and these 900 proteins are primarily of eukaryotic origin (table S6).

Imported *B. bigelowii*–encoded proteins completed some missing steps in biosynthetic pathways that are incomplete and only partially encoded in the UCYN-A genome. For example, the UCYN-A genome lacks genes required for biosynthesis of certain amino acids, nucleotides, vitamins, and cofactors (*16*, *24*). *B. bigelowii* proteins predicted to localize to UCYN-A (by differential protein abundance or the presence of a uTP sequence) were able to complete the pathway for biosynthesis of threonine and contribute to the incomplete proline and serine biosynthetic pathways (Fig. 3A and data S3 and S7). The UCYN-A genome is also missing orotate phosphoribosyltransferase (PyrE), which is required for de novo biosynthesis of pyrimidine nucleotides (*33*). *B. bigelowii*–encoded PyrE was found to be enriched in UCYN-A samples and contained a uTP (Fig. 3A). Both UCYN-A and *B. bigelowii* encode dihydroorotase (PyrC) required for pyrimidine biosynthesis, and both were enriched in UCYN-A and present in similar quantities. The bifunctional enzyme dihydrofolate reductase–thymidylate synthase (DHFR-TS) is targeted to UCYN-A with a uTP, and it is the only enzyme missing from the tetrahydrofolate biosynthetic pathway (Fig. 3A). These findings suggest that the evolutionary reduction of the UCYN-A genome was likely influenced by redundancies in the *B. bigelowii* genome.

Many of the most abundant *B. bigelowii* proteins found in UCYN-A are known to support N_2 fixation in other diazotrophs. For instance, flavodoxin is the preferred electron donor to nitrogenase in a free-living relative of UCYN-A (*34*), and both ferredoxin and flavodoxin are known electron donors to nitrogenase (*35*, *36*). However, the UCYN-A genome lacks flavodoxin altogether, which has led to the assumption that ferredoxin is responsible for nitrogenase reduction (*21*, *24*). A *B. bigelowii*–encoded flavodoxin (FldA) containing a uTP was among the most common *B. bigelowii* proteins in UCYN-A during the day (Fig. 3B and data S4 and S5). UCYN-A contains five ferredoxins, which were detected in our proteomes (data S5). Two of these were more abundant in UCYN-A during the day, including a 2Fe-2S plant type and a 4Fe-4S Nif-specific type. The remaining three UCYN-A–encoded ferredoxins were among the few UCYN-A proteins with higher abundances in the whole-cell samples, indicating export from UCYN-A to *B. bigelowii*. Ferredoxin is known to cross membranes by means of specific transporters (*37*), but we detected no homologs of these proteins in either *B. bigelowii* or UCYN-A. The dynamics of these electron donors and their redox partners are yet unclear, but it appears that redox activity is coordinated between UCYN-A and *B. bigelowii*, and it is possible that *B. bigelowii*–encoded flavodoxin is an electron donor to UCYN-A nitrogenase.

Diazotrophs are particularly susceptible to damage by reactive oxygen and N species produced by high respiratory rates in the vicinity of nitrogenase and consequently use antioxidant defenses (*38*). Several *B. bigelowii* proteins involved in mediating oxidative damage were found in UCYN-A. Cytochrome P450s (CYPs) are heme-thiolate–containing oxygenases that serve diverse functional roles (*39*), and certain cytochrome P450s are involved in the biosynthesis of antioxidant flavonoids (*40*). Out of 24 *B. bigelowii* CYP protein sequences, 10 contained a uTP sequence including all five CYP proteins detected by mass spectrometry. All of these were more abundant in UCYN-A samples, and four of the five detected CYPs were more abundant during the day, when N_2 fixation is occurring (Fig. 3B and data S4). All uTP-containing CYPs are class I cytochrome P450s predicted to use a ferredoxin redox partner (*41*). *B. bigelowii*–encoded glutathione dehydrogenase, which is involved in the glutathione antioxidant response (*42*), was detected to be more abundant in UCYN-A samples and contains a uTP (data S4). Also abundant in UCYN-A was a *B. bigelowii*–encoded heme oxygenase (HemeO), which could be participating in heme catabolism for Fe acquisition or in the antioxidant defense of nitrogenase (*34*, *43*). Diazotrophs that are actively fixing N_2 also have elevated levels of biochemical reactions fueled by respiration and that involve strong reducing potentials (*38*). The data show that in UCYN-A, this activity appears to also be accompanied by import of protective *B. bigelowii* proteins (Fig. 3B).

Other imported proteins are potentially involved in regulation of UCYN-A by the *B. bigelowii* cell. Methionine-tRNA formyltransferase (Fmt) is critical for initiating translation in prokaryotes, mitochondria, and plastids of eukaryotes through import of a nuclear-encoded protein (*44*). The *fmt* gene is absent from the UCYN-A genome, but a *B. bigelowii*–encoded Fmt with a uTP is present in the UCYN-A proteome, particularly during the day (Fig. 3B and data S6).

The elongation factor Tu (EF-Tu) is a highly abundant bacterial protein essential for protein synthesis in prokaryotes but is also known to localize to the cell surface, where its functions are related to pathogenesis (*45*). *B. bigelowii* encodes an EF-Tu protein likely horizontally acquired from Alphaproteobacteria, which was very highly abundant in UCYN-A. This protein does not contain a uTP sequence, which may be an indication that it adheres to the outer membrane rather than being imported. UCYN-A also encodes its own EF-Tu, which was also highly abundant in the UCYN-A proteome. Bacterial EF-Tu proteins are known to be shared between different species (*46*) and to bind proteins of their symbiotic partners (*47*). This protein in UCYN-A likely mediates interactions with the surrounding *B. bigelowii* cell (Fig. 3B and data S6).

As previously discussed (Fig. 1), the UCYN-A light-dark and division cycles are important for maintaining symbiosis. Circadian rhythms are important in many cyanobacteria for daily cycles and cell division. UCYN-A lacks two of the three *kai* genes (*16*, *48*), which are responsible for generation of circadian rhythms in other cyanobacteria (*49*). In algae and higher plants, nuclear-encoded cryptochrome (Cry) proteins are known to regulate chloroplast gene expression and coordinate circadian cycles across organelles (*50*). Two *B. bigelowii*–encoded Cry proteins, both containing uTP sequences, were highly abundant in UCYN-A (Fig. 2B and data S6). Their presence in UCYN-A is a further indication of *B. bigelowii* control over fundamental UCYN-A diel processes such as cell cycle progression.

UCYN-A thus has a mosaic proteome made up of proteins encoded by its own genome and those from the *B. bigelowii* genome that contain likely signal sequences for targeting and transport into UCYN-A. Many of these proteins are isofunctional with proteins lost during reduction of the UCYN-A genome, and some are redundant with existing UCYN-A proteins. The essential nature of some of these proteins suggests that control over translation and replication has been transferred to *B. bigelowii*, which also appears to perform a substantial portion of UCYN-A protein synthesis. Most of these proteins are retargeted eukaryotic proteins, as is the case in the *Paulinella* chromatophore (*32*, *51*), which is consistent with the interpretation that both are in an early stage of organellogenesis arising from a similar duration of endosymbiosis (table S7) (*52*, *53*).

UCYN-A as a nitrogen-fixing organelle

The relationship between UCYN-A and *B. bigelowii* has been assumed to be symbiotic since its initial discovery (*8*, *11*, *54*). This study shows that UCYN-A has evolved beyond endosymbiosis and has characteristics of a N_2-fixing organelle, or "nitroplast," an organelle that has previously been hypothesized (*51*). Such an organelle would be an N_2-fixing analog of primary carbon-fixing plastids, which appear to have only evolved twice, the first being the cyanobacterial endosymbiont ancestor of the Archaeplastida (*4*, *29*), and the second being the photosynthetic organelle of the protist *Paulinella* resulting from the endosymbiosis between *Paulinella* and a cyanobacterium related to *Synechococcus* (*55*). Genome reduction is a characteristic of both endosymbiosis and organellogenesis (*56*). There are now clear parallels among evolution of mitochondria, chloroplasts, the chromatophore, and UCYN-A with respect to genome reduction, protein import, and size relationship to the cell. The *Paulinella* chromatophore and UCYN-A appear to both be in an intermediate stage of organellogenesis and have undergone gene loss resulting in genomes of ~1 to 1.5 Mb. However, UCYN-A has undergone more-extensive gene loss, as N_2-fixing unicellular cyanobacteria typically have genomes approaching 5 Mb, compared with 3 Mb for the non–N_2-fixing cyanobacteria (such as *Synechococcus*), which are the proposed ancestors of the *Paulinella* chromatophore, and ~3 Mb for the diatom spheroid bodies (*51*). UCYN-A also appears to be more efficient, with doubling times of ~1.4 days (fig. S3C), compared with 5 to 7 days for the chromatophore (*4*). Thus, UCYN-A might be further along the spectrum of organellogenesis than the *Paulinella* chromatophore (*4*) or the diatom spheroid bodies (*51*).

Model systems that provide a more complete picture of the events that occur along the path of organellogenesis are exceedingly rare. Although it has been suggested that there is no reason why an N_2-fixing organelle could not be maintained in the cell, such organelles have yet to be described (*51*). The findings of this study, along with the regulation of spheroid body size (*17*) and coevolution of *B. bigelowii* and UCYN-A (*53*), suggest that the evolution of the nitroplast is underway and establish *B. bigelowii* as an important model system of great interest for understanding organellogenesis, with implications for contemporary efforts to engineer N_2-fixing capabilities into agricultural plants (*10*, *57*).

REFERENCES AND NOTES

1. S. Paracer, V. Ahmadjian, *Symbiosis: An Introduction to Biological Associations* (Oxford Univ. Press, ed. 2, 2000).
2. J. M. Archibald, *Curr. Biol.* **25**, R911–R921 (2015).
3. G. Attardi, G. Schatz, *Annu. Rev. Cell Biol.* **4**, 289–333 (1988).
4. T. G. Stephens, A. Gabr, V. Calatrava, A. R. Grossman, D. Bhattacharya, *New Phytol.* **231**, 1693–1699 (2021).
5. C. Kneip, P. Lockhart, C. Voss, U.-G. Maier, *BMC Evol. Biol.* **7**, 55 (2007).
6. R. A. Foster, J. P. Zehr, *Annu. Rev. Microbiol.* **73**, 435–456 (2019).
7. A. Caputo, J. A. A. Nylander, R. A. Foster, *FEMS Microbiol. Lett.* **366**, fny297 (2019).
8. A. W. Thompson *et al.*, *Science* **337**, 1546–1550 (2012).
9. T. Coba de la Peña, E. Fedorova, J. J. Pueyo, M. M. Lucas, *Front. Plant Sci.* **8**, 2229 (2018).
10. J. Elhai, Engineering of crop plants to facilitate bottom-up innovation: A possible role for broad host-range nitroplasts and neoplasts. OSF Preprints (2023); https://doi.org/10.31219/osf.io/ny2rc.
11. K. Hagino, R. Onuma, M. Kawachi, T. Horiguchi, *PLOS ONE* **8**, e81749 (2013).
12. S. Suzuki *et al.*, *Front. Plant Sci.* **12**, 749895 (2021).
13. K. A. Turk-Kubo, H. M. Farnelid, I. N. Shilova, B. Henke, J. P. Zehr, *J. Phycol.* **53**, 451–461 (2017).

14. K. Harding *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 13371–13375 (2018).
15. P. H. Moisander *et al*., *Science* **327**, 1512–1514 (2010).
16. H. J. Tripp *et al*., *Nature* **464**, 90–94 (2010).
17. F. M. Cornejo-Castillo, K. Inomura, J. P. Zehr, M. J. Follows, *Cell* **187**, 1762–1768.e9 (2024).
18. A. A. Ekman *et al*., in *Handbook of X-ray Imaging: Physics and Technology*, P. Russo, Ed. (CRC Press, 2017), pp. 817–852.
19. P. J. Keeling, J. M. Archibald, *Curr. Biol.* **18**, R345–R347 (2008).
20. J. P. Zehr *et al*., *Science* **322**, 1110–1112 (2008).
21. D. Sarkar *et al*., *PLOS Comput. Biol.* **17**, e1008983 (2021).
22. M. R. Gradoville *et al*., *Environ. Microbiol.* **23**, 4518–4531 (2021).
23. R. L. Robson, J. R. Postgate, *Annu. Rev. Microbiol.* **34**, 183–207 (1980).
24. M. D. C. Muñoz-Marín *et al*., *mBio* **10**, e02495-18 (2019).
25. C. A. Larabell, K. A. Nugent, *Curr. Opin. Struct. Biol.* **20**, 623–631 (2010).
26. V. Loconte *et al*., *Structure* **30**, 510–521.e3 (2022).
27. M. A. Le Gros *et al*., *Cell Rep.* **17**, 2125–2136 (2016).
28. A. E. Douglas, J. A. Raven, *Philos. Trans. R. Soc. London Ser. B* **358**, 5–18 (2003).
29. S. D. Dyall, M. T. Brown, P. J. Johnson, *Science* **304**, 253–257 (2004).
30. B. D. Bruce, *Trends Cell Biol.* **10**, 440–447 (2000).
31. K. G. Hansen, J. M. Herrmann, *Protein J.* **38**, 330–342 (2019).
32. A. Singer *et al*., *Curr. Biol.* **27**, 2763–2773.e5 (2017).
33. K. F. Jensen, *J. Bacteriol.* **175**, 3401–3407 (1993).
34. M. A. Saito *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2184–2189 (2011).
35. L. E. Mortenson, *Annu. Rev. Microbiol.* **17**, 115–138 (1963).
36. M. G. Yates, *FEBS Lett.* **27**, 63–67 (1972).
37. M. Wojnowska, D. Walker, *mBio* **11**, e02081-20 (2020).
38. M. Becana, M. A. Matamoros, M. Udvardi, D. A. Dalton, *New Phytol.* **188**, 960–976 (2010).
39. I. G. Denisov, T. M. Makris, S. G. Sligar, I. Schlichting, *Chem. Rev.* **105**, 2253–2277 (2005).
40. B. A. Pandian, R. Sathishraj, M. Djanaguiraman, P. V. V. Prasad, M. Jugulam, *Antioxidants* **9**, 454 (2020).
41. S. Li, L. Du, R. Bernhardt, *Trends Microbiol.* **28**, 445–454 (2020).
42. M. Ramírez *et al*., *Physiol. Plant.* **149**, 389–407 (2013).
43. Y. Zhou *et al*., *New Phytol.* **239**, 1989–2006 (2023).
44. B. S. Laursen, H. P. Sørensen, K. K. Mortensen, H. U. Sperling-Petersen, *Microbiol. Mol. Biol. Rev.* **69**, 101–123 (2005).
45. K. L. Harvey, V. M. Jarocki, I. G. Charles, S. P. Djordjevic, *Front. Microbiol.* **10**, 2351 (2019).
46. K. Nishiyama *et al*., *Appl. Environ. Microbiol.* **89**, e02190-22 (2023).
47. M. Widjaja *et al*., *Sci. Rep.* **7**, 11227 (2017).
48. D. Bombar, P. Heller, P. Sanchez-Baracaldo, B. J. Carter, J. P. Zehr, *ISME J.* **8**, 2530–2542 (2014).
49. T. Kondo, M. Ishiura, *BioEssays* **22**, 10–15 (2000).
50. J. H. C. Griffin, K. Prado, P. Sutton, G. Toledo-Ortiz, *Physiol. Plant.* **169**, 515–528 (2020).
51. E. C. M. Nowack, A. P. M. Weber, *Annu. Rev. Plant Biol.* **69**, 51–84 (2018).
52. L. Delaye, C. Valadez-Cano, B. Pérez-Zamorano, *PLOS Curr.* **8**, ecurrents.tol.e68a099364bb1a1e129a17b4e06b0c6b (2016).
53. F. M. Cornejo-Castillo *et al*., *Nat. Commun.* **7**, 11071 (2016).
54. J. P. Zehr, I. N. Shilova, H. M. Farnelid, M. d. C. Muñoz-Marín, K. A. Turk-Kubo, *Nat. Microbiol.* **2**, 16214 (2017).
55. H. S. Yoon, A. Reyes-Prieto, M. Melkonian, D. Bhattacharya, *Curr. Biol.* **16**, R670–R672 (2006).
56. N. A. Moran, *Cell* **108**, 583–586 (2002).
57. K. Mahmud, S. Makaju, R. Ibrahim, A. Missaoui, *Plants* **9**, 97 (2020).
58. Y. Perez-Riverol *et al*., *Nucleic Acids Res.* **50**, D543–D552 (2022).
59. T. H. Coale *et al*., Data for ADK1075 - Nitrogen fixing organelle in a marine alga, Dryad (2024); https://doi.org/10.5061/dryad.2z34tmptf.

ACKNOWLEDGMENTS

This work is dedicated to the memory of Kenneth H. Coale and his love for science. Thanks to G. Grigorean for performing liquid chromatography–tandem mass spectrometry and data analysis in the Proteomics Core Facility of the Genome Center, University of California, Davis. We thank S. Suzuki for providing assembled *B. bigelowii* transcripts. We acknowledge K. Marchus for performing carbon and nitrogen measurements in the Analytical Lab at the University of California, Santa Barbara. **Funding:** This work was supported by the following: Simons Foundation grants 724220 and 824082 (J.P.Z.); NIH NIGMS P30GM138441 (C.L.; The National Center for X-ray Tomography); DOE's Office of Biological and Environmental Research DE-AC02-05CH11231 (C.L.; The National Center for X-ray Tomography); and JSPS KAKENHI grants 17K05694 and 20H04325 (K.H.). **Author contributions:** Conceptualization: T.H.C., V.L., K.A.T.-K., C.L., and J.P.Z. Methodology: T.H.C., V.L., K.A.T.-K., W.K.E.M., S.C., K.H., C.L., and J.P.Z. Strain isolation: K.H., Y.T., T.N., and M.A. SXT investigation: V.L., K.A.T.-K., B.V., A.E., J.H.-C., M.L.G., and C.L. Proteomics investigation: T.H.C. Visualization: T.H.C., V.L., and C.L. Funding acquisition: K.H., C.L., and J.P.Z. Project administration: C.L. and J.P.Z. Supervision: C.L. and J.P.Z. Writing – original draft: T.H.C., V.L., K.A.T.-K., C.L., and J.P.Z. Writing – review & editing: T.H.C., V.L., K.A.T.-K., W.K.E.M., S.C., K.H., M.L.G., C.L., and J.P.Z. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** *B. bigelowii* (FR-21) plastid sequencing data used in this paper has been deposited in GenBank under the following accession numbers: OR912953, OR912954, and OR912955. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (*58*) partner repository with the dataset identifier PXD047956. Datasets S1 to S7, quantitative proteomics data, protein sequence data, and SXT data are available in Dryad (*59*). **License information:** Copyright © 2024 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk1075
Materials and Methods
Figs. S1 to S6
Tables S1 to S7
References (*60*–*91*)
MDAR Reproducibility Checklist
Movies S1 to S3

Submitted 19 December 2023; accepted 22 February 2024
10.1126/science.adk1075